DOCUMENT RESUME

ED 435 730 TM 030 362

AUTHOR Ediger, Marlow
TITLE Statistics, the Teacher, and the Principal.
PUB DATE 1999-00-00
NOTE 9p.
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Administrators; Communication (Thought Transfer); Criterion Referenced Tests; *Educational Research; Elementary Secondary Education; Norm Referenced Tests; Principals; Reliability; *Research Reports; School Statistics; *Statistics; Teachers; *Test Results; Validity

ABSTRACT

Teachers and principals need to understand and be able to talk intelligently about statistics. They need to be able to clarify concepts and generalizations used in tests that measure student progress. Among the terms educators need to be able to explain are "mean," "average," "median," "validity," "reliability," "pilot studies," and "item analysis." Educators also need to be able to communicate the meanings of percentiles and standard deviations and the distinction between standardized and criterion referenced tests. Teachers, administrators, and measurement specialists should develop a booklet of major statistical concepts and generalizations to inform parents and others as to the meaning of these terms. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 730

STATISTICS, THE TEACHER, AND THE PRINCIPAL

Marlow Ediger

TM030362

Marlow Ediger

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

STATISTICS, THE TEACHER, AND THE PRINCIPAL

There is an important need for teachers and administrators to understand and be able to converse intelligently about statistics. The knowledge and skills involving statistics need to be used by teachers and administrators to clarify with others what is involved in statewide and national testing. Certainly, the testing and measurement movement is here and very much alive. Can teachers and administrators clarify needed concepts and generalizations in tests used to measure pupil progress?

What Do Teachers And Principals Need To Know?

When conducting parent/teacher conferences and in talking with student teachers/ cooperating teachers, whom I supervise in the public schools, I have identified selected concepts and generalizations, which will now be discussed.

Educators in the public schools need to understand and be able to communicate meanings pertaining to the concept "mean" (1). Averages can be very misleading when, for example, looking at the average salaries of teachers in the United States. If principals teach one class or a half-hour a day, they may be counted in the group of teachers. If the principal then makes $65,000 a year, the average salary of teachers goes up much. A better term to use here would be "median." The median is the middle most salary of teachers, arranged sequentially from high to low. This lops off the extremely high and extremely low salaries. I agree strongly with the statement, "I am sitting on a slab of ice and have my feet in boiling water, but the temperature reading is "average" between the two."

In addition to needing to understand the concepts of "average" and "median," school personnel should also understand and communicate to parents, concepts relating to testing such as "validity," "reliability," "pilot studies," and "item analysis." Why? Unless these concepts are implemented, the state-mandated test may have little value in terms of pupil results when testing

time comes. "Validity" means the test actually measures what it says it measures, and not something else. Thus, a mathematics test truly measurers learner achievement in mathematics, not in some other academic area. Validity, too, emphasizes measuring what pupils have had opportunities to learn. Pupils may not have had opportunities to learn subject matter on a test. "Reliability" means the test measures consistently. Thus, if a pupil would take the same test the second time, results would be the same/similar. Otherwise, the test has little value, if the two measures differ much from each other.

With "pilot studies," weaknesses are taken out as pupils in the pilot study respond to the test items, prior to it being given to pupils statewide. Not always do different states conduct pilot studies of their tests prior to implementation, whereby all pupils are to take these tests within their borders (2).

Percentiles And Standard Deviation

Test results from individual pupils may be given in terms of percentiles. Thus, for example, a pupil is on the fiftieth percentile based on results from test taking. This means that for everyone one hundred pupils having taken the same test in the pilot study, fifty were above and fifty were below the fiftieth percentile. Very frequently, a school or state desires to have all of their pupils achieve above the fiftieth percentile. Is this possible? Theoretically, no it is not possible. The median or middle most score of test takers is the fiftieth percentile. Students then varied in test taking results from the ninety-ninth to the first percentile. With all pupils achieving above the fiftieth percentile, the norms in the pilot study would need to be changed so the middle most score would again be the fiftieth percentile. The range again being from the ninety-ninth to the first percentile.

What about the standard deviation from the mean (or average score) from those taking the test in the pilot study? Many times, parents and other lay people, believe all pupils should achieve

above the mean or average for good teaching to have been in evidence. Is this possible? The answer is a definite "no." Why? If all pupils achieve above the average, then the norms of the test in the pilot study need to be revised so that half the test takers will be above and half below the mean. The mean pertains, always, to the average score of all test takers and of those in the pilot study. If all test takers receive scores above the mean, then a new mean will need to be determined since the mean refers to the average scores of test takers (3).

Standard deviations are given in terms of one, two, or three standard deviations, either above or below the mean. Thus, 34% of the test takers will be one standard deviation (SD) above or below the mean. Thirteen percent will be two SDs either above or below the mean whereas, approximately two and one-half percent of pupils will be either 3 SDs above and below the mean (or average). I talked with a school principal who was working towards all pupils in a school being one SD or higher above the mean in test results. Good luck! However, the norms of the test will then need redoing so that the mean of all scores of the test takers will be the average. Then, results of test takers can be reordered in terms of one, two, or three SDs above or below the mean.

Standardized Versus Criterion Preferred Tests

Standardized tests, developed and published by a commercial company, emphasize the following:

1. Every pupil takes the same achievement test.
2. Each pupil is given the same amount of time to take the test.
3. The directions given to take the test are the same for all test takers.
4. There are right and wrong answers to be given by pupils for each test item.
5. Multiple choice items are usually used in the standardized test.
6. Pupil's results may be machine scored in a matter of seconds, based on a correct scoring key.

7. Results given are in terms of percentiles and standard deviations.

8. Rank order of pupils in a class may be given or determined readily by the teacher as to who is first, second, third high, and so on in the class. Rank order is very important to notice in that the results from pupils may come very close together as to who is first, second, third, and so on in the class (too close to call).

9. Comparisons are made among and between pupils, in terms of test results.

10. Test items have been selected and written under the auspices of the commercial company which publishes the standardized test given in the school, where pupils are being tested.

Are the ten above named standards appropriate to use in completing the writing and publishing of the standardized test? Not exactly. There are major loopholes here when providing for pupils of diverse achievement levels. For every action, there seemingly is an opposite and equal reaction. I will list these reactions for a few of the numbered items, given above, as standards on which standardized tests were based.

1 & 2. Pupils can now differ much from each other in achievement. Perhaps, it is all right for learners to take the same test items, if they are valid. Validity is weak in most standardized tests since the test items are not aligned with specific objectives emphasized in the classroom. Apparently, no standardized test has accompanying objectives for teachers to emphasize in teaching. A major problem enters in when all pupils receive the same amount of time when responding to the test. Learners, individually, certainly can respond at different rates of speed to test items. On a teacher written test, each pupil should have the necessary time when responding to the total test in order to complete the test satisfactorily.

3 & 4. A few pupils need more time than others when absorbing information for direction taking. This may not be a serious problem if the test directions are clearly and concisely given. The

direction provider needs to be certain that each pupil understands how to take the standardized test, when the sample items are given and pupils individually mark each for the examiner to notice.

Right and wrong answers do present a problem in answers to multiple choice items. It is difficult to write quality multiple choice items that have four plausible responses of same/similar length. A pupil might well raise questions as to the meaning of one or more responses. At the same time, the teacher is not to give test information to the pupil which leads to the correct response. If responses are too precise, factual learning is being tested, rather than critical and creative thinking, as well as problem solving (4).

Criterion referenced tests (CRTs) attempt to take out selected weaknesses of standardized tests. CRTs, developed on the state level under the auspices of the state department of education, do have specific objectives for teachers to use in teaching pupils. These objectives provide a benchmark as to what to teach. The test items of the CRT generally are closely aligned with these objectives. Thus, items of the CRT should be much more valid than those on the standardized test.

Multiple choice items are predominate on the CRT. Pupils are required to be tested on selected grade levels, as mandated by the state. How well pupils do on the CRT might be compared on a report card, school district by school district comparisons. Teachers are to be held accountable in terms of how well pupils have done on the test. There are a few states that have educational bankruptcy laws in which a school district can be taken over by the state if test results are continually low.

As compared to standardized test results, CRTs will not tend to have the following:

1. A range of scores from the ninety-ninth to the first percentile. Why? Ideally, many pupils are to achieve the specific objectives and do well on the CRT.
2. Low validity. Thus, the teacher may teach to having pupils achieve the specific objectives, provided by the state. Test items are aligned with the stated objectives.

3. Low reliability with test/retest; split half or alternative forms reliability. However, CRTs may have lower reliability if not pilot tested to take out weak test items (5).

In Conclusion

Teachers, administrators, and measurement specialists need to develop a booklet of major statistical concepts and generalizations to inform parents and others as to the meaning of these terms. Statistical data can certainly be misinterpreted and/or misused. For example, when listing pupils in rank order from obtained test results, the differences among individuals can be quite small from the highest listed results from a pupil, compared to several below. Actually, the differences might be so small that they are miniscule. In other words, the scores may be so close from these pupils that the differences do not matter.

As a further example, in experimental studies, the mean differences in achievement between the two groups may not be statistically significant at the .05 or .01 level, and yet the differences are great when visually and intellectually examining the differences between the experimental and control groups in a research design.

Selected References

1. Ediger, Marlow (1997), Teaching Reading and the Language Arts in the Elementary School. Kirksville, Missouri: Simpson Publishing Company, Chapter Five.

2. Ediger, Marlow (1995), Philosophy in Curriculum Development. Kirksville, Missouri: Simpson Publishing Company, Chapter Six.

3. Ediger, Marlow (1997), The Modern Elementary School. Kirksville, Missouri: Simpson Publishing Company, Chapter Eight.

4. Hopkins, Kenneth D., and Julian C. Stanley (1991), Educational and Psychological Measurement and Evaluation, Sixth Edition. Englewood Cliffs, New Jersey, Chapter Three.

5. Walsh, W. Bruce, and Nancy E. Betz (1985), Tests and Assessment. Englewood Cliffs, New Jersey, Chapter Four.